Archives of Suicide Research

ISSN: 1381-1118 (Print) 1543-6136 (Online) Journal homepage: https://www.tandfonline.com/loi/usui20

Using Machine Learning to Identify Suicide Risk: A Classification Tree Approach to Prospectively Identify Adolescent Suicide Attempters

Ryan M. Hill, Benjamin Oosterhoff & Calvin Do

Archives of Suicide Research, 24:218–235, 2020

ISSN: 1381-1118 print/1543-6136 online
DOI: 10.1080/13811118.2019.1615018

Routledge
Taylor & Francis Group

Check for updates

Using Machine Learning to Identify Suicide Risk: A Classification Tree Approach to Prospectively Identify Adolescent Suicide Attempters

Ryan M. Hill, Benjamin Oosterhoff, and Calvin Do

This study applies classification tree analysis to prospectively identify suicide attempters among a large adolescent community sample, to demonstrate the strengths and limitations of this approach for risk identification. Data were drawn from the National Longitudinal Study of Adolescent to Adult Health. Youth ($n = 4,834$, $M_{age} = 16.15$, $SD = 1.63$, 52.3% female, 63.7% White) completed at-home interviews at Wave 1 and a measure of suicide attempts 12 months later, at Wave 2. Results indicated two classification tree solutions that maximized risk prediction, with 69.8%/85.7% sensitivity/specificity and 90.6%/70.9% sensitivity/specificity, respectively. Classification trees provide a technique for identification of individuals at-risk for suicide attempts. Classification trees produce easy-to-implement decision rules and tailored screening approaches that can be adapted to the goals of a particular organization.

Keywords adolescent, classification tree analysis, machine learning, suicide attempt

Suicide is the second-leading cause of death for individuals between the ages of 13–35 years in the United States (Centers for Disease Control and Prevention, 2019). Suicide attempts are also frequent during adolescence: Data from the nationally representative Youth Risk Behavior Survey indicate that 8.6% of high school students made a suicide attempt in the previous 12 months (Kann et al., 2016). Critically, suicide attempts are a robust predictor of future suicidal behavior (Brown, Beck, Steer, & Grisham, 2000; Nordstrom, Asberg, Aberg-Wistedt, & Nordin, 1995) and the economic impact of suicide and suicide attempts among 15–24-year-olds was estimated at over $15.5 billion in 2013 alone (Shepard, Gurewich, Lwin, Reed, & Silverman, 2016). Hence, the identification of at-risk youth is critical to reducing both the rate of suicide-related behaviors among adolescents and the associated economic and mental health burden.

A large number of risk markers for suicide attempts have been identified in the empirical literature. Even so, few evidence-based approaches have been developed to assist clinicians in the process of integrating this wide array of risk markers into an assessment of individual risk. Clinical risk assessments typically evaluate recent suicidal cognitions and history of suicidal behaviors, to inform immediate responses, such as the need for inpatient hospitalization to ensure patient safety (e.g., Osman et al., 2001; Posner et al., 2011). Although some guidance exists for evaluating immediate suicide risk (e.g., Chu et al., 2015), prediction of suicide attempts beyond the proximal period is necessary to allow sufficient time to administer preventive interventions. To date, longer-term prospective identification of suicide-related behaviors has proven challenging (Franklin et al., 2016; Walsh, Ribeiro, & Franklin, 2017).

Providers assessing long-term suicide risk must apply knowledge of a vast empirical literature to a complex array of risk and protective factors. With little guidance regarding how heavily to weigh individual markers of risk, providers must rely on personal experience and clinical intuition when evaluating an individual clients' risk for suicide. Indeed, clinical decision making may not be optimally suited for evaluating complex patterns of risk, as it is subject to influence from a variety of personal factors, including provider age, religiosity, and relationship status, and is subject to substantial variability across providers (Berman, Stark, Cooperman, Wilhelm, & Cohen, 2015; Regehr, LeBlanc, Bobo, Paterson, & Birze, 2015, for discussions of clinical decision making, see Bryan & Rudd, 2006; Joiner, Walker, Rudd, & Jobes, 1999). It is understandable, then, that empirical evaluations of clinical decision making have shown relatively little accuracy in prospectively predicting suicide attempts (Wang et al., 2016).

To help address the limitations of clinical decision making, researchers have focused on empirical approaches to prospectively identify suicide risk. These empirical approaches have historically been dominated by regression-based models, which have demonstrated accuracy little better than clinical-judgments (e.g., Bolton, Spiwak, & Sareen, 2012; Goldstein, Black, Nasrallah, & Winokur, 1991; for an exception, see McCarthy et al., 2015). However, regression models often fail to account for dynamic interactions among risk markers, in which a given risk marker is only predictive (or its predictive value is of greater magnitude) for certain subsets of youth. Although interactions may be applied within regression-based frameworks, they are specified *a priori* which typically leads researchers to investigate a limited array of possible interactive combinations (Tabachnick & Fidell, 2007). Thus, regression-based models may not account for the complex interplay of risk markers associated with suicide-related behaviors.

Machine learning approaches provide a promising set of techniques for improving predictive modeling of suicide-related behaviors (e.g., Hill, Oosterhoff, & Kaplow, 2017; Kessler et al., 2017; Mann et al., 2008; Walsh, Ribeiro, & Franklin, 2017). Machine learning broadly refers to the ability of computer systems to progressively improve statistical prediction of a given outcome without being explicitly programed (Murphy, 2012). Scholars often apply machine learning to explore complex predictive solutions using large data repositories, such as genetic databases (e.g., Goldberg & Holland, 1988), functional imagining studies (e.g., Just et al., 2017; Richiardi, Achard, Bunke, & Van De Ville, 2013), or nationally representative survey data (e.g., Hill, Oosterhoff, &

Kaplow, 2017). One key advantage of machine learning approaches is the ability to better account for large numbers of inter-related predictors and to examine all possible interactions between predictors. As a result, these techniques may be better able to model the complex interplay of suicide risk markers in a way that regression-based models and human cognition often do not.

Recent research demonstrates the promise of machine learning approaches for predicting suicide-related behaviors: For example, a machine learning technique known as random forest modeling correctly identified 3,188 suicide attempts up to 720 days in advance, while missing only 62 suicide attempts among a sample of 5,543 high-risk adult patients with a history of self-injury (Walsh, Ribeiro, & Franklin, 2017). This model identified fewer than 900 false positives, defined as cases in which a subject was incorrectly identified as a suicide attempter (Walsh, Ribeiro, & Franklin, 2017). Similarly, using data from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS), a machine learning model identified that 45.6% of suicide deaths occurring within 26 weeks of outpatient mental health visits were among the 15% of visits deemed to be at highest risk (Kessler et al., 2017). Classification tree analysis, another type of machine learning, was used to prospectively identify adolescents with suicide ideation over a one-year period, using data from the National Longitudinal Study of Adolescent to Adult Health (Hill, Oosterhoff, & Kaplow, 2017). That study was able to provide multiple classification models for use in various clinical settings, correctly identifying between 44.6% and 77.6% of suicide ideators (Hill, Oosterhoff, & Kaplow, 2017). These applications show the potential of machine learning techniques for enhancing the identification of individuals at-risk for suicide sufficiently far in advance to provide a variety of preventive interventions.

AN INTRODUCTION TO CLASSIFICATION TREE ANALYSIS

The present study applies classification tree analysis (CTA), a branch of machine learning, to prospectively identify suicide attempters among a large community sample of adolescents assessed at two time points, approximately one year apart. CTA may be an especially useful application of machine learning to the field of suicide prevention because it (a) holds promise to improve prediction of suicide risk while (b) providing a clinically applicable decision tree that outlines rules used to determine whether someone is at risk for a suicide attempt. More accurately identifying youth at risk for suicide attempts in a prospective fashion may help enhance suicide prevention efforts by providing a longer window of opportunity for the application of preventive interventions.

CTA produces a set of decision rules for categorizing subjects into identified outcome groups. To create this set of decision rules, CTA identifies the best predictor of the outcome, partitions data into subgroups based on that predictor, then repeats the process with each subgroup (for an introduction to CTA see, Breiman, Friedman, Stone, & Olshen, 1984). The result is a classification tree that describes how to apply decision rules based on the modeled predictors to new cases to determine the expected outcome for a given case. The resulting classification tree provides important advantages over other machine learning techniques (e.g., neural networks) in the context of predicting suicide attempts. Unlike other forms of machine learning, the classification tree produced by the CTA provides a transparent, accessible process for achieving a given solution that can reproduced and integrated using standard survey methods within clinical settings. Importantly, this transparency is also necessary to further theory regarding the complex interactions among risk factors for suicide.

A simple example of a classification tree is provided in Figure 1, in which a fictional sample of 1,000 youth have been screened for elevated depressive symptoms. In this example, the top "node" of the tree provides the overall sample size and the incidence rate (i.e., the percentage of the sample for which the target outcome is present; 20.0%). This node then "branches" into two smaller nodes based on the variable "*age,*" with each node containing a subset of the original data set. The leftmost node contains $n = 600$ youth ages 13 years and below, of whom 10% reported elevated depressive symptoms. The rightmost node contains the $n = 400$ youth over the age of 13 years, of whom 35% reported elevated depressive symptoms. This node splits again, based on the variable "*sex,*" with females ($n = 250$) having a higher incidence rate (30%) than males ($n = 250$; 10%). As indicated by bold font, the classification tree would predict that females above age 13 will report elevated depressive symptoms, whereas males above age 13 and youth below age 13 will not. In more complex classification trees, these branches may have additional splits.

CTA operates by first building a classification tree, then "pruning" the resulting tree, to balance accuracy and replicability. The procedures by which CTA approaches the partitioning of data and prunes the

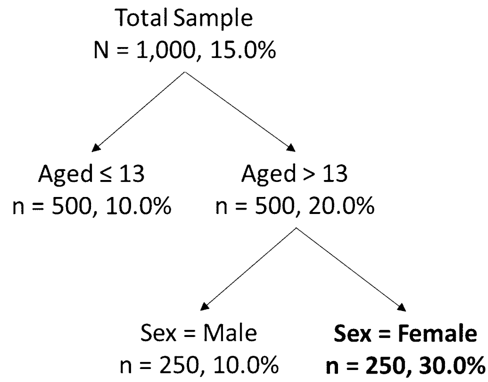

FIGURE 1. *Sample classification tree. Note. Figure presents a fictional example (not based on actual data) illustrating classification tree analysis of 1000 youth screened for elevated depressive symptoms.*

resulting tree vary depending on the program used and specifications selected. As applied here, CTA first identifies the optimal branching variable by examining all possible variables (and pairs of variables) to select the best branch. Then the optimal cut point is determined for that variable. (For a discussion of variable selection algorithms and the determination of split points, see Kim & Loh, 2001, 2003). The data then branch along that decision point into two nodes. This procedure is repeated within each node, and additional branches are made. This continues until no additional significant improvements in classification are possible based on the variables provided. Alternatively, branching stops when a minimum sample size is reached within a node, at which point further divisions of that node are not evaluated.

This CTA procedure emphasizes classification accuracy and produces an overly complex tree, which is "overfitted" to the available data (De'ath & Fabricius, 2000; Kim & Loh, 2001). An overfitted tree has too many branches, which may capitalize upon unique characteristics of the data. Because these characteristics are less likely to replicate in other samples, these extra branches may limit the accuracy of the tree when applied to new cases. To manage the problem of overfitting, the classification tree is "pruned," a process in which branches are removed to produce a smaller tree. (For a discussion of classification tree pruning procedures, see De'ath & Fabricius, 2000).

To optimize categorization of subjects into outcome groups, each classification tree is designed to minimize the "misclassification cost" of the overall tree. Misclassification cost refers to a penalty for mistakenly classifying an individual as a member of any group other than the one to which they belong. Two types of misclassification are possible. The first are "missed cases" or "false negatives" in which the case is classified as negative for the outcome when, in fact, the outcome did occur. Here, missed cases represent suicide attempters that the classification tree labeled as nonattempters, failing to identify them as at-risk. The second type of misclassifications are "false positives," in which a case is classified as positive for the outcome when the outcome did not occur. Here, false positives are youth identified as suicide attempters who did not report a suicide attempt.

Critically, the real-world implications of these two misclassification errors are not always equal. For suicide prevention, it is generally agreed that the real-world cost of missing an at-risk youth is greater than the cost of potentially delivering preventive interventions to a "false positive." Fortunately, misclassification cost can be modified to include greater penalties for specific misclassifications. As is done here, a greater penalty for missed cases can be applied in comparison to false positives. This places greater emphasis on reducing missed cases at the cost of introducing false positives. Increasing the cost for missed cases has the effect of increasing sensitivity (correct identification of suicide attempters) at the expense of specificity (correct identification of nonattempters).

In the following investigation, we will provide a demonstration of a classification tree approach to prospectively identifying suicide attempters among an adolescent sample. From a methodological standpoint, we apply CTA to elucidate the strengths and limitations of this branch of machine learning and to describe how CTA differs from traditional approaches to risk identification. This study seeks to emphasize the unique insights to be gleaned from classification tree approaches and highlight the potential applications of classification tree models. (For readers interested in the more technical nuances of CTA, we refer them to Breiman et al., 1984; De'ath & Fabricius, 2000; Kim & Loh, 2001, 2003).

METHOD

Participants

The current sample was drawn from the National Longitudinal Study of Adolescent to Adult Health (Add Health) database. The data was obtained through the Inter-University Consortium for Political and Social Science Research (ICPSR; Harris & Udry, 2016). Add Health participants provided written informed consent for participation in all aspects of Add Health and the study was approved by the appropriate Institutional Review Board. In 1994–1995, Add Health representatives surveyed youth attending a nationally stratified sample of high schools and middle schools. As part of the overall study, a subset of participants from these high schools completed at-home interviews. In Wave 1 (1994–1995), 20,745 adolescents (M_{age} = 15.0 years, range 11–21 years) completed the first assessment involving in-home interviews using computer-assisted interview methods. A sample of $N = 14{,}738$ participants were then re-interviewed one year after the initial assessment (Wave 2). This study utilizes the public use dataset, which contains a random subsample drawn from the larger Add Health database. The public use data included responses for $n = 6{,}054$ youth at Wave 1, and $n = 4{,}834$ youth reassessed at Wave 2.

Consistent with guidelines for Add Health data users (Chantala, 2006), we limited our analyses to $n = 4{,}834$ (M_{age} = 16.15, $SD = 1.63$ at Wave 1) youth who completed both waves of at-home interviews. Participants (52.3% female) were primarily White (63.7%), African American/Black (23.7%), Asian (3.7%), Native American (3.1%), and other (6.1%). The median household income was $40,000 (range: $0 – $900,000). Regarding parents' education, 14.9% of parent respondents (15.8% of spouses) did not complete high school, 40.9% completed high school but did not have any college training (38.0% spouses), 19.2% completed some college (15.9% spouses), and 25.1% obtained a college degree or higher (29.7% of spouses).

Measures

The Add Health study was designed to examine adolescent's social, economic, psychological, and physical well-being within the context of family, peers, neighborhoods, communities, schools, and romantic relationships. Thus, the adolescent self-report data from the Add Health study covered a wide variety of psychological, behavioral, and environmental constructs relevant for adolescent suicide risk, including mental health, violence exposure, negative life events, and community engagement, as well as family, peer, and school functioning (see Harris, 2013).

For the present study, the Add Health Wave 1 data codebook was reviewed. Adolescent self-report variables were considered for inclusion. Variables were excluded if more than 20% of data were missing or if the variable was deemed uninterpretable by the research team. All other variables were included in the analysis. Responses labeled *missing, refused, don't know, not applicable,* and *legitimate skip* were considered missing for the purpose of this analysis. Of note, the response *legitimate skip* indicated that the respondent was not asked a question due to a previous response indicating the question was not applicable. For example, youth not enrolled in school in the current/most recent school year were not asked questions about school attendance and grades and *legitimate skip* was recorded for these items. Due to the theoretical importance of, and overwhelming empirical support for, history of suicide attempts as a predictor of future suicide-related behaviors (e.g., Gould, Greenberg, Velting, & Shaffer, 2003; Joiner, 2005; O'Connor, 2011), the item *"During the past 12 months, how many times did you actually attempt suicide?"* was retained, with *legitimate skip* recoded as *0* (no suicide attempts). This procedure resulted in 345 variables included in the CTA. A full list of variables included in the analysis is available in the supplemental files; Table 1 provides the wording and response options for all variables appearing in the final solutions.

TABLE 1. **Variables Appearing in Classification Trees**

Suicide-Related Variables

During the past 12 months, did you ever seriously ***think about*** committing ***suicide***? (H1SU1)
0 = No; 1 = Yes

Have any of your ***friends*** have succeeded in committing ***suicide*** in the past 12 months? (H1SU5)
0 = No; 1 = Yes

Depression-Related/Physiological Variables

During the past 12 months, how often have you ***felt very tired*** for no reason? (H1GH8) 0 = Never; 1 = Just a few times; 2 = About once a week; 3 = Almost every day; 4 = Every day

In the past 12 months, how often have you ***been moody***? (H1GH20) 0 = Never; 1 = Just a few times; 2 = About once a week; 3 = Almost every day; 4 = Every day

In the past 12 months, how often have you ***cried frequently*** (H1GH21) 0 = Never; 1 = Just a few times; 2 = About once a week; 3 = Almost every day; 4 = Every day

During the past 12 months, how often have you had ***cold sweats***? (H1GH5) 0 = Never; 1 = Just a few times; 2 = About once a week; 3 = Almost every day; 4 = Every day

Substance Use Variables

During the past 30 days, on how many days did you use ***chewing tobacco***? (H1TO10) Numerical response from 0 to 30.

How old were you when you first tried any other type of ***illegal drug***, such as LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills, without a doctor's prescription? (H1TO40) 0 = Never, 1 = Any reported age

HIV/STD/Sex-Related Variables

How many ***people*** do ***you know*** who ***have AIDS***? (H1GH45) Numerical response from 0 to 99.

What do you think your ***chances*** are ***of*** getting another ***sexually transmitted disease***, such as gonorrhea or genital herpes? (H1GH46) 1 = Very high; 2 = High; 3 = Low; 4 = Very low; 5 = No chance

Have you ever given someone ***sex in exchange for drugs or money***? (H1NR3) 0 = No; 1 = Yes

What do you think the ***chances*** are that the following will happen to you? You will get ***HIV or AIDS***. (H1EE15) 1 = Almost no chance; 2 = Some chance, but probably not; 3 = A 50-50 chance; 4 = A good chance; 5 = Almost certain

In the past year, have you received ***testing or treatmentfor a sexually transmitted disease*** or AIDS? (H1HS9) 0 = No; 1 = Yes

In the past year have you ***receivedfamily planning*** counseling or services? (H1HS7) 0 = No; 1 = Yes

Risky Behavior Variables

How often did you ***drive a car without*** the owner's ***permission***? (H1DS8) 0 = Never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times

In the past 12 months, how often did you ***paint graffiti*** or signs on someone else's property or in a public place? (H1DS1) 0 = Never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times

How often did you ***run away from home***? (H1DS7) 0 = Never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times

During the past 30 days, on how many days did you ***carry a weapon***–such as a gun, knife, or club–to school? (H1FV9) 0 = None; 1 = 1 day; 2 = 2 or 3 days; 3 = 4 or 5 days; 4 = 6 or more days

(*Continued*)

Relationship Variables

Please imagine your ideal romantic relationship in the next year. I would ***meet my partner's parents***. (H1ID1B) 1 = Card kept; 2 = Card rejected

Please imagine your ideal romantic relationship in the next year. We would ***think of ourselves as a couple***. (H1ID1K) 1 = Card kept; 2 = Card rejected

Resident Mother Variables

How ***far in school*** did she [resident ***mother***] go? (H1RM1) 1 = 8th grade or less; 2 = More than 8th grade, but did not graduate from high school; 3 = Went to a business, trade, or vocation school instead of high school; 4 = High school graduate; 5 = Completed a GED;6 = Went to a business, trade, or vocation school after high school; 7 = Went to college, but did not graduate;8 = Graduated from a college or university; 9 = Professional training beyond a four year college or university;10 = She never went to school;11= She went to school, but respondent doesn't know what level;12 = Respondent doesn't know if she went to school

What ***kind of work*** does she [resident ***mother***] do? (H1RM4) 1 = Professional, such as doctor, lawyer, scientist; 2 = Professional, such as teacher, librarian, or nurse; 3 = Manager; 4 = Technical; 5 = Office worker; 6 = Sales worker; 7 = Restaurant worker or personal service; 8 = Craftsperson; 9 = Construction worker; 10 = Mechanic; 11 = Factory worker or laborer; 12 = Transportation; 13 = Military or security; 14 = Farm or fishery worker; 15 = Other; 16 = None

Demographics

What is your race: ***Asian or Pacific Islander.*** (H1GI6D) 0 = No; 1 = Yes

Note. Bold/italicized text indicates text abbreviation used in classification tree figures; parentheses indicate variable name in Add Health data set.

The dependent variable was the presence of a suicide attempt in the previous 12 months, assessed at Wave 2, by the item *"During the past 12 months, how many times did you actually attempt suicide?"* For this item, as for Wave 1, *legitimate skip* responses ($n = 4311$) were recoded as *0*, resulting in $n = 4{,}642$ nonattempters. Responses indicative of 1 or more suicide attempts, including *1 time* ($n = 116$), *2 or 3 times* ($n = 52$), *4 or 5 times* ($n = 14$), *and 6 or more times* ($n = 10$), were recoded as *1,* resulting in 192 suicide attempters.

Data Analysis

CTA were conducted using CRUISE version 3.6.4 software (Kim & Loh, 2008). The CRUISE software is freely available from the developers along with documentation (http://www.stat.wisc.edu/~loh/cruise.html). Independent variables, as described above, were coded as categorical variables if they included nominal responses (e.g., *yes/no, male/*female). Independent variables that included Likert-type scales or continuous responses were coded as numerical variables (e.g., age in years, scales ranging from *strongly agree* to *strongly disagree*).

CTA were conducted according to the recommendations of Kim and Loh (2003, 2008). Classification trees were generated via univariate splits using the 2 D variable selection method with splits selected via discriminant analysis. A minimum node size of 24 (representing approximately 0.5% of the sample) was specified, such that nodes containing fewer than 24 cases were not evaluated for further branches. Missing values were addressed by fitting all available cases prior to nodewise mean/mode imputation to allow cases with missing values to branch accordingly. Prior class probabilities were estimated based on the data. Misclassification costs were sequentially increased from 1:1 (i.e., equal "cost" or penalty for misclassification for false negatives as compared with false positives) to 62:1 (i.e., a 62 times greater penalty for false negatives as compared with false positives) in order to generate classification trees with sensitivity values ranging from 0 to 1. Pruning was conducted via 10-fold cross-validation and a 0.5 SE selection rule. Examples of the CRUISE input files and the CRUISE commands selected for the present analysis are available in the supplemental material. The utility of each classification tree was determined based on the sensitivity and specificity, as well as favoring a low ratio of false positives to true positives.

RESULTS

A series of CTAs were generated based on misclassification cost ratios ranging from 1:1 to 62:1, producing 26 unique classification trees. Descriptive information is provided for each tree in Table 2, including overall classification accuracy, the number of nodes in the classification tree, the number of suicide attempters correctly identified and missed, and the number of false positives. The sensitivity and specificity of each solution was plotted (Figure 2). As seen in Table 2, Trees 1–4 had relatively poor sensitivity, with fewer than 30% of suicide attempters identified. As seen in Figure 2, Trees 5–9 cluster together with regard to specificity (83.7% to 89.0%), though they varied in sensitivity (55.7% to 71.9%). Visual inspection of the trees themselves reveals that the Trees in this cluster share a similar pattern of initial branches, with branches further along the tree differing and providing variation in sensitivity and specificity. Among these, Tree 7 appears to provide the greatest overall sensitivity and specificity, representing the point furthest from the diagonal in Figure 2. Similarly, Trees 10–15 and 17–26 also appear to cluster together in terms of specificity (66.9%-77.1%), again with varying sensitivity (76.6%-90.6%). As before, these trees share a similar pattern of initial branches, with branches further along the tree differing. Among this cluster, Tree 15 appears to provide the greatest overall sensitivity and specificity. Trees 16, 24, and 25 show similar sensitivity to the previous cluster, but with noticeably poorer specificity. Finally, Tree 17 has 0 specificity, identifying all adolescents as suicide attempters. Thus, Trees 7 and 15 appear to offer solutions that maximize improvement over chance classification.

Tree 7. Moderate Sensitivity, Moderate Specificity

Tree 7, depicted in Figure 3, represents a classification tree using a 22:1 misclassification cost ratio. Tree 7 identified 16.5% of adolescents as suicide attempters, correctly identifying 134 of 192 suicide attempters and an additional 633 false positives. For Tree 7, the ratio of false positives to true positives was 4.95:1, with an overall accuracy rate of 85.1%. That is, for every correctly identified suicide attempter, 4.95 adolescents were incorrectly classified as suicide attempters.

Tree 7 can be interpreted as follows: The initial sample contained 4,834 individuals, of whom 4.0% reported a suicide attempt at Wave 2. The first split was a history of suicide ideation at Wave 1. Of the 618 adolescents with a history of

TABLE 2. **Descriptive Information for Classification Trees**

Tree	Cost Ratio	Terminal Nodes	Percent Correct	True Positives	True Negatives	False Positives	False Negatives	False Positive: True Positive	Sensitivity (%)	Specificity (%)
1	1:1	1	96.03	0	4642	0	192	—	0.00	100.00
2	2:1	4	96.28	29	4625	17	163	0.59:1	15.10	99.63
3	3:1	4	95.74	41	4587	55	151	1.34:1	21.35	99.82
4	4:1-6:1	3	94.70	57	4521	121	135	2.12:1	29.69	97.39
5	7:1-19:1, 21:1	2	87.67	107	4131	511	85	4.78:1	55.73	88.99
6	20:1, 25:1-30:1	4	83.12	128	3890	752	64	5.88:1	66.67	83.80
7	**22:1**	**8**	**85.08**	**134**	**3979**	**663**	**58**	**4.95:1**	**69.79**	**85.72**
8	23:1	3	86.59	115	4071	571	77	4.97:1	59.90	87.70
9	24:1	5	83.20	138	3884	758	54	5.49:1	71.88	83.67
10	31:1-33:1	7	77.06	147	3578	1064	45	7.24:1	76.56	77.08
11	34:1	13	74.29	160	3431	1211	32	7.57:1	83.33	73.91
12	35:1-36:1	10	75.05	155	3473	1169	37	7.54:1	80.73	74.82
13	37:1-38:1	14	76.56	157	3544	1098	35	6.99:1	81.77	76.35
14	39:1	16	71.06	169	3266	1376	23	8.14:1	88.02	70.36
15	**40:1-41:1**	**21**	**71.68**	**174**	**3291**	**1351**	**18**	**7.76:1**	**90.63**	**70.90**
16	42:1	4	49.77	170	2236	2406	22	14.15:1	88.54	48.17
17	43:1	17	72.96	166	3361	1281	26	7.72:1	86.46	72.40
18	44:1, 46:1	20	72.38	171	3328	1314	21	7.68:1	89.06	71.69
19	45:1	6	67.38	153	3104	1538	39	10.05:1	79.69	66.87
20	47:1-48:1	20	72.30	171	3324	1318	21	7.71:1	89.06	71.61
21	49:1-50:1, 52:1	21	71.87	171	3303	1339	21	7.83:1	89.06	71.15
22	51:1	20	72.76	169	3348	1294	23	7.66:1	88.02	72.12
23	53:1-54:1	20	72.47	169	3334	1308	23	7.74:1	88.02	71.82
24	55:1-57:1	15	63.28	171	2888	1754	21	10.26:1	89.06	62.21
25	58:1-61:1	17	61.98	175	2821	1821	17	10.41:1	91.15	60.77
26	62:1	1	3.97	192	0	4642	0	24.18:1	100.00	0.00

Note. Bold indicates trees presented in the text. Cost ratios indicate *cost for false negative: cost of false positive*.

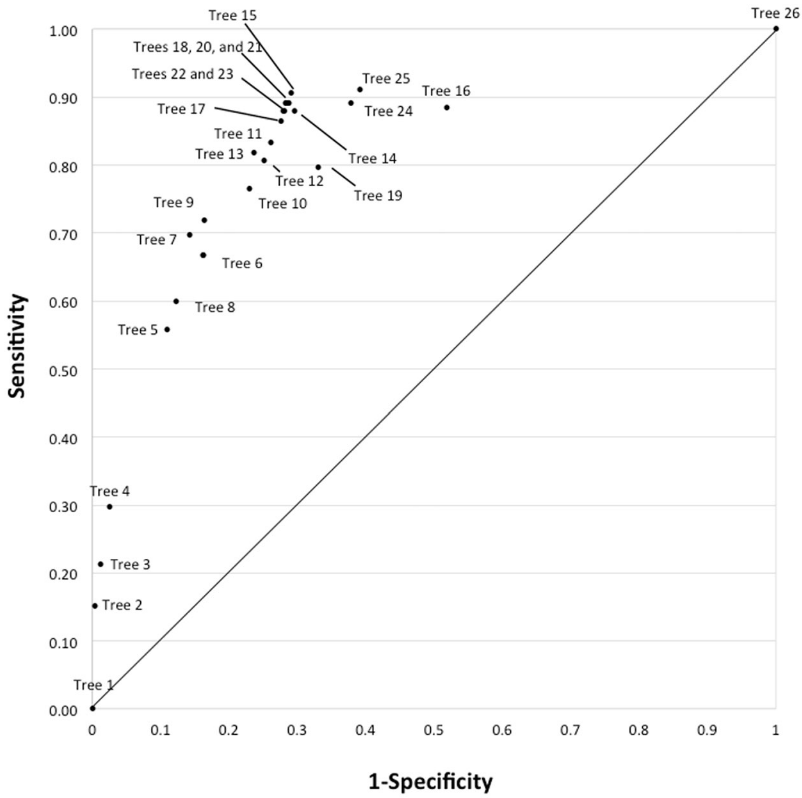

FIGURE 2. *Sensitivity and specificity of classification trees.*

suicide ideation at Wave 1, 17.3% were suicide attempters at Wave 2, compared with only 2.0% of the 4,216 adolescents without a history of suicide ideation at Wave 1. Among adolescents without a history of suicide ideation at Wave 1, subsequent splits were based on frequency of feeling tired for no reason, perceived chance of getting a sexually transmitted disease, running away from home, Asian or Pacific Islander ethnicity, and having a friend die by suicide in the previous year.

Tree 15. High Sensitivity, Low Specificity

Tree 15, depicted in Figure 4, represents classification trees using 40:1 and 41:1 misclassification cost ratios, as they produced identical trees. Tree 15 identified 31.5% of adolescents as suicide attempters, correctly identifying 174 of 192 suicide attempters and an additional 1,351 false positives. For Tree 15, the ratio of false positives to true positives was 7.76:1, with an overall accuracy rate of 71.7%.

As with Tree 7, the initial split in Tree 15 is based on adolescents' history of suicide ideation at Wave 1. However, Tree 15 is substantially more complex, including a total of 14 different variables. Additional splits in Tree 15 fell, broadly, into common categories, including symptoms of depression and physiological symptoms (feeling tired for no reason,

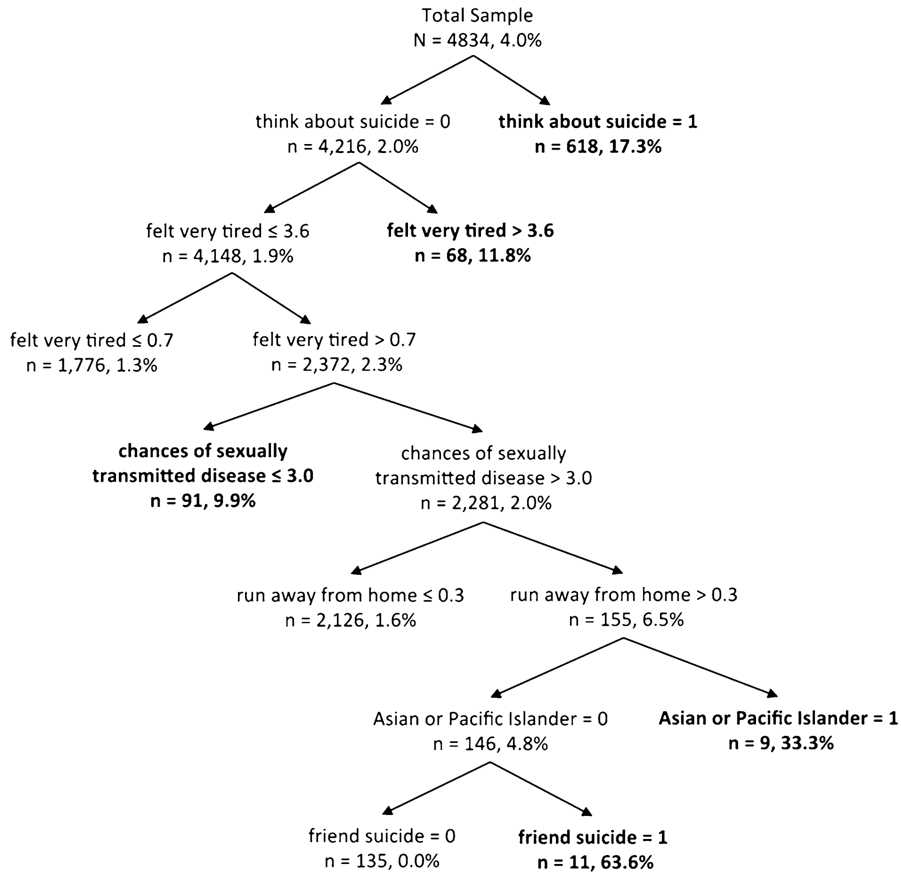

FIGURE 3. *Classification Tree 7. Note. Bold text indicates cases classified as suicide attempters; n = sample size at that node; percentages are incidence rates. See Table 2 for exact item wording.*

being moody, crying frequently, and frequency of having cold sweats), familial characteristics (mothers education and type of work), indicators of risky behaviors (frequency of painting graffiti, driving a car without the owner's permission, and carrying a weapon), sex and sexually transmitted disease-related variables (perceived chance of contracting HIV or AIDS, receiving testing for sexually transmitted diseases, receiving family planning services, exchanging sex for drugs or money, knowing people with AIDS), substance use (trying illegal drugs, use of chewing tobacco), and expected characteristics of a romantic relationship (see Figure 4 for the progression of these splits).

DISCUSSION

The current study aimed to demonstrate the utility of CTA, a branch of machine learning, for improving efforts to identify youth at-risk for suicide-related behaviors. Data from the National Longitudinal Study of Adolescent to Adult Health was utilized to identify adolescents who

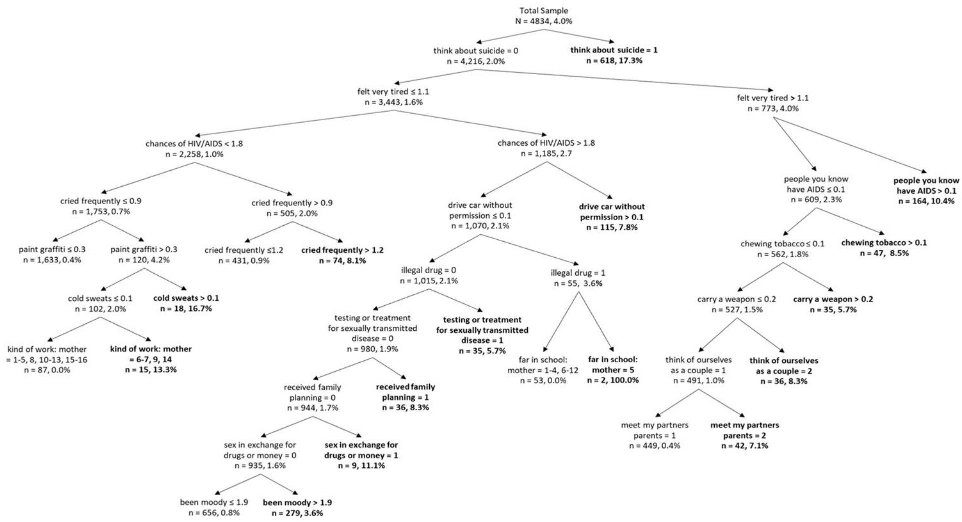

FIGURE 4. *Classification Tree 15. Note. Bold text indicates cases classified as suicide attempters; n = sample size at that node; percentages are incidence rates. See Table 2 for exact item wording.*

reported a suicide attempt at Wave 2 of the study, based on responses provided one year prior. Results indicated that Trees 7 and 15 provided distinct solutions that maximized improvements over chance classifications. These two classification trees varied in complexity, sensitivity, specificity, and thus provide organizations with flexibility in selecting a screening approach based on the resources available for providing preventative interventions.

The selection of an appropriate classification tree depends on the goals and resources of the organization implementing the screening, as well as the sensitivity and specificity of the tree. A tree with high sensitivity accurately identifies a large percentage of suicide attempters but may do so at the cost of specificity, resulting in a higher rate of false positives. For an organization providing low-cost services, high sensitivity, even at the cost of reduced specificity, may be optimal. For organizations providing services with greater intensity, specificity may be given greater priority, in order to reduce the overall demand for services.

Tree 7 showed moderate sensitivity, identifying almost 70% of suicide attempters, with a moderate number of false positives. As a screening tool, Tree 7 would identify 16.5% of adolescents as being at-risk, or approximately 1 in 6 youth. For a complete screening, Tree 7 would require a maximum of 7 unique pieces of information. Tree 7 would be suitable for organizations with limited resources or that wish to provide brief individualized services. Though the majority of suicide attempters would receive intervention, substantial resources would be directed to false positives.

Tree 15 correctly identified over 90% of suicide attempters. It provides a highly sensitive solution, but at the cost of reduced specificity, with nearly 8 false positives for every correctly identified suicide attempter. The high rate of false positives resulted in a tree that identifies 31.5% of youth as being at-risk, or nearly 1 in 3 youth screened. Although Tree 15 correctly identifies the vast majority of suicide attempters, the need to provide services to 1 in 3 youth may strain the resources of many organizations. Thus, Tree 15 would be more suitable in a setting in which a positive screen results in the provision of psychoeducational resources or brief technology-focused interventions (e.g., a web-based prevention program; Hill & Pettit, 2019). For a complete screening, Tree 15 would require a maximum of 9 unique pieces of information.

Advantages and Disadvantages of CTA

CTA has several advantages that hold promise for enhancing research and practice regarding suicide prevention. The machine learning nature of CTA facilitates the use of large-scale data, which may be especially important as technological growth and open data policies are providing greater access to "big data" sources. Further, the ability to use large data to test complex interactions among risk markers holds promise to ultimately increase the accuracy of models seeking to predict suicide risk relative to regression-based approaches. A unique advantage of CTA is the capability to produce clinically interpretable and meaningful decision rules for risk-screening that can be tailored for specific communities, organizations, and subpopulations. The flexibility in specifying misclassification costs allows individuals to generate predictive solutions with varying levels of complexity, sensitivity, and specificity, which provides further adaptation of solutions for organizations with varying resources (see Hill, Oosterhoff, & Kaplow, 2017).

Classification trees provide clinically useful decision-making tools, as depicted in Figures 3 and 4, while minimizing the number of points of information needed to reach a decision. Individual screening questions are determined based on previous splits, until a terminal node is reached. For example, for an adolescent reporting suicide ideation, no further information is needed and a designation of at-risk is reached with only a single point of information. For Tree 15, a maximum of nine pieces of information are needed, despite the seeming complexity of the tree. Thus, CTA not only holds promise for improving predictive models of suicide risk, but also produces tangible decision rules that can be flexibly tailored and easily applied with a variety of different organizational settings.

This analysis utilized a 12-month follow-up period to demonstrate a model of early risk identification that can support implementation of a broad range of preventive interventions, from early or "upstream" prevention programs (e.g., Hill & Pettit, 2019) to more acute-risk focused interventions (e.g., Stanley and Brown, 2012). A 12-month follow-up period is also consistent with the schedule of pediatric primary care well-visits recommended by the American Academy of Pediatrics (Committee on Practice and Ambulatory Medicine, AAP Bright Futures Periodicity Schedule Workgroup, 2019). However, it is worth noting that prediction of suicide attempts over longer follow-up periods is challenging (Franklin et al., 2016; Walsh, Ribeiro, & Franklin, 2017). Selection of a longer follow-up period likely also impacts selected predictors, shifting the focus away from acute risk indicators (e.g., suicide ideation severity, suicide planning) and toward distal risk indicators (e.g., depressive symptoms, behavioral problems, and interpersonal difficulties).

Despite the advantages of CTA, there are also some notable limitations that need to be considered when applying this procedure for suicide prevention. First, CTA is an exploratory data analytic approach and thus heavily relies on the *quality* of the data used in the model. If theoretically relevant constructs are missing from data sources or if the data suffers from other limitations (e.g., high degree of missingness), the predictive algorithm and accompanying decision tree may have a high degree of bias. This is an important limitation to consider, especially given that many secondary data sources were not specifically designed to identify suicide risk. For example, the present study examined data from the National Longitudinal Study of Adolescent to Adult Health, which included only a single item indicator of suicide attempts which was only administered to those with suicide ideation. Single-item indicators may overestimate actual suicide attempts, a limitation of the models presented here (Hom, Joiner, & Bernert, 2016; Millner, Lee, & Nock, 2015). Additionally, the data was collected between 1994 and 1996, prior to a 30% increase in the rate of suicide in the United States, which occurred between 1999 and 2017 (Centers for Disease Control and Prevention, 2019). This highlights the need for replication of this study with more recent data.

The exploratory nature and dependency on available data may also pose challenges for theory building and integration of findings across studies, which may produce drastically different results. Lastly, the application of CTA requires several specifications and accompanying assumptions. Although the use of CTA within the field of risk prevention is growing, few guidelines are available regarding appropriate sample sizes or model specification selections given the specific questions or data structures.

Future Directions

CTA, and machine learning techniques in general, offer an opportunity to examine data sources in a new way and from a different methodological perspective. However, CTA's data-driven approach and interaction-based framework is not easily interpreted in terms of theory development and model building. Thus, although CTA may provide an informative method for examining clinical risk identification, it will be important to recognize the strengths and limitations of CTA and to utilize CTA as one statistical tool among many. Drawing theoretical interpretations and generalizable knowledge from CTA should be done with caution and will require identification of patterns of results across multiple models and in multiple settings. Similarly, due to the potential for overfitting classification trees, it will be critically important to replicate classification trees across data sets, particularly before implementing them as large-scale screening tools.

The implementation of classification trees based on large data sets may also raise ethical concerns regarding the use of patient or personal data (Tucker, Tackett, Glickman, & Reger, 2019). Large data sets are often the product of governmental records, electronic medical records, and school records, or may derive from social media or other large data repositories. The use of this data, particularly for developing and implementing classification tree-based screening tools, may introduce new clinical and ethical concerns. For example, physician obligations for acting on distal suicide risk are not clearly outlined and classification trees using existing medical record data may introduce concerns regarding patient consent for use of protected health information. (For a discussion of potential ethical issues in the use of large data repositories for suicide prevention efforts, see Tucker et al., 2019.)

Even so, as the quality of big data sources improves, CTA may offer an increasing accurate method for predicting suicide-related behaviors. The importance of accurate screening and risk-identification approaches, particularly those that do not rely solely or explicitly on self-identification of suicide risk (Just et al., 2017), is difficult to understate. The dissemination of selective and indicated suicide prevention approaches are critically dependent on the accurate identification of youth at-risk for suicide related behaviors. If CTA and other machine learning techniques offer new insights into risk identification, it is crucial that they be explored.

AUTHOR NOTE

Ryan M. Hill, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA.

Benjamin Oosterhoff, Department of Psychology, Montana State University, Bozeman, MT, USA.

Calvin Do, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA.

Correspondence concerning this article should be addressed to Ryan M. Hill, Ph.D., Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, 1102 Bates Ave., Suite C.0235.05A, Houston, TX 77030, USA. E-mail: Ryan.Hill@bcm.edu

DISCLOSURE STATEMENT

The authors declare no conflicts of interest.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Berman, N. C., Stark, A., Cooperman, A., Wilhelm, S., & Cohen, I. G. (2015). Effect of patient and therapist factors on suicide risk assessment. *Death Studies*, *39*(7), 433–441. doi: 10.1080/07481187.2014.958630

Bolton, J. M., Spiwak, R., & Sareen, J. (2012). Predicting suicide attempts with the SAD PERSONS scale: A longitudinal analysis. *The Journal of Clinical Psychiatry*, *73*(3), e735–e741. doi:10.4088/JCP.11m07362

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks.

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. *Journal of Consulting and Clinical Psychology*, *68*(3), 371–377. doi:10.1037//0022-006.X.68.3.371

Bryan, C. J., & Rudd, M. D. (2006). Advances in the assessment of suicide risk. *Journal of Clinical Psychology*, *62*(2), 185–200. doi:10.1002/jclp.20222

Centers for Disease Control and Prevention. (2019). *Injury prevention & control: Data & statistics (WISQARS)*. http://www.cdc.gov/injury/wisqars/index.html. Accessed 20 March 2019.

Chantala, K. (2006). *Guidelines for analyzing add health data*. Chapel Hill, NC: Carolina Population Center: University of North Carolina Chapel Hill.

Chu, C., Klein, K. M., Buchman-Schmitt, J. M., Hom, M. A., Hagan, C. R., & Joiner, T. E. (2015). Routinized assessment of suicide risk in clinical practice: An empirically informed

update. *Journal of Clinical Psychology*, *71*(12), 1186–1200. doi:10.1002/jclp.22210
Committee on Practice and Ambulatory Medicine, AAP Bright Futures Periodicity Schedule Workgroup. (2019). 2019 recommendations for preventive pediatric health care. *Pediatrics*, *143*(3).
De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, *81*(11), 3178–3192. doi:10.2307/177409
Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., … Nock, M. K. (2016). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 197–232.
Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, *3*(2/3), 95–99.
Goldstein, R. B., Black, D. W., Nasrallah, A., & Winokur, G. (1991). The prediction of suicide: Sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. *Archives of General Psychiatry*, *48*(5), 418–422.
Gould, M. S., Greenberg, T., Velting, D. M., & Shaffer, D. (2003). Youth suicide risk and preventive interventions: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(4), 386–405. doi:10.1097/01.CHI.0000046821.95464.CF
Harris, K. M. (2013). *The Add Health Study: Design and accomplishments.* Retrieved from http://www.cpc.unc.edu/projects/addhealth/data/guides/DesignPaperWIIV.pdf
Harris, K. M., & Udry, J. R. (2016). National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use]. ICPSR21600-v17. Chapel Hill, NC: Carolina Population Center, University of North Carolina-Chapel Hill/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 05–25. http://doi.org/10.3886/ICPSR21600.v17
Hill, R. M., Oosterhoff, B., & Kaplow, J. B. (2017). Prospective identification of adolescent suicidal ideation using classification tree analysis: Models for community-based screening. *Journal of Consulting and Clinical Psychology*, *85*(7), 702–711. doi:10.1037/ccp0000218
Hill, R. M., & Pettit, J. W. (2019). Pilot randomized controlled trial of LEAP: A program to reduce adolescent perceived burdensomeness. *Journal of Clinical Child and Adolescent Psychology*, *48*(sup1), S45–S56. doi:10.1080/15374416.2016.1188705
Hom, M. A., Joiner, T. E., & Bernert, R. A. (2016). Limitations of a single-item assessment of suicide attempt history: Implications for standardized suicide risk assessment. *Psychological Assessment*, *28*(8), 1026–1030. doi:10.1037/pas0000241
Joiner, T. E. (2005). *Why people die by suicide*. Cambridge, MA, US: Harvard University Press.
Joiner, T. E., Walker, R. L., Rudd, M. D., & Jobes, D. A. (1999). Scientizing and routinizing the assessment of suicidality in outpatient practice. *Professional Psychology: Research and Practice*, *30*(5), 447–453. doi:10.1037//0735-7028.30.5.447
Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D., Cha, C., Nock, M. K., & Brent, D. (2017). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour*, *1*, 911–919. doi:10.1038/s41562-017-0234-y
Kann, L., McManus, T., Harris, W. A., Shanklin, S. L., Flint, K. H., Hawkins, J., … Zaza, S. (2016). Youth risk behavior surveillance – United States, 2015. *MMWR. Surveillance Summaries*, *65*(6), 1–174. doi:10.15585/mmwr.ss6506a1
Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., … Ursano, R. J. (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Molecular Psychiatry*, *22*(4), 544–551. doi:10.1038/mp.2016.110
Kim, H., & Loh, W. (2001). Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, *96*, 598–604.
Kim, H., & Loh, W. (2003). Classification trees with bivariate linear discriminant node models. *Journal of Computational and Graphical Statistics*, *12*(3), 512–530. doi:10.1198/1061860032049
Kim, H., & Loh, W. (2008). CRUISE user manual. *Department of Statistics, University of Wisconsin*, Madison. Technical Report 989.

Mann, J. J., Ellis, S. P., Waternaux, C. M., Liu, X., Oquendo, M. A., Malone, K. M., … Currier, D. (2008). Classification trees distinguish suicide attempters in major psychiatric disorders: A model of clinical decision-making. *The Journal of Clinical Psychiatry*, *69*(1), 23–31. doi:10.4088/JCP.v69n0104

McCarthy, J. F., Bossarte, R. M., Katz, I. R., Thompson, C., Kemp, J., Hannemann, C. M., … Schoenbaum, M. (2015). Predictive modeling and concentration of the risk of suicide: Implications for preventive intervention in the US Department of Veterans Affairs. *American Journal of Public Health*, *105*(9), 1935–1942. doi:10.2105/AJPH.2015.302737

Millner, A. J., Lee, M. D., & Nock, M. K. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLoS ONE*, *10*(10), e0141606. doi:10.1371/jounal.pone.0141606

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT press.

Nordstrom, P., Asberg, M., Aberg-Wistedt, A., & Nordin, C. (1995). Attempted suicide predicts suicide risk in mood disorders. *Acta Psychiatrica Scandinavica*, *92*(5), 345–350. doi:10.1111/j.1600-0447.1995.tb.09595.x

O'Connor, R. C. (2011). The integrated motivational-volitional model of suicidal behavior. *Crisis*, *32*(6), 295–298. doi:10.1027/0337-5910/a000120

Osman, A., Bagge, C. L., Guitierrez, P. M., Konick, L. C., Kooper, B. A., & Barrios, F. X. (2001). The Suicidal Behaviors Questionnaire-Revised (SBQ-R): Validation with clinical and nonclinical samples. *Assessment*, *2001*(5), 443–454. doi:10.1177/1073191101008 00409

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., … Mann, J. J. (2011). The Columbia–Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*, *168*(12), 1266–1277. doi:10.1176/appi.ajp.2 011.10111704

Regehr, C., LeBlanc, V. R., Bogo, M., Paterson, J., & Birze, A. (2015). Suicide risk assessments: Examining influences on clinicians' professional judgement. *American Journal of Orthopsychiatry*, *85*(4), 295–301. doi:10.1037/ort0000075

Richiardi, J., Achard, S., Bunke, H., & Van De Ville, D. (2013). Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing Magazine*, *30*(3), 58–70. doi:10.1109/MSP.2012.2233865

Shepard, D. S., Gurewich, D., Lwin, A. K., Reed, G. A., & Silverman, M. M. (2016). Suicide and suicidal attempts in the United States: Costs and policy implications. *Suicide and Life-Threatening Behavior*, *46*(3), 352–362. doi:10.1111/sltb.12225

Stanley, B., & Brown, G.K. (2012). Safety Planning Intervention: A Brief Intervention to Mitigate Suicide Risk. *Cognitive and Behavioral Practice*, *19*(2), 256–264. doi:10.1016/j.cbpra.2011.01.001

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Tucker, R.P., Tackett, M.J., Glickman, D., & Reger, M.A. (2019). Ethical and practical considerations in the use of a predictive model to trigger suicide prevention interventions in healthcare settings. *Suicide and Life-Threatening Behavior*, *49*(2), 382–392. doi:10.1111/sltb.12431

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457–469. doi:10.1177/2167702617691560

Wang, Y., Bhaskaran, J., Sareen, J., Bolton, S.-L., Chateau, D., & Bolton, J. M. (2016). Clinician prediction of future suicide attempts: A longitudinal study. *The Canadian Journal of Psychiatry*, *61*(7), 428–432. doi:10.1177/0706743716645287